DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

Generalized Portfolio Performance Measures: Optimal Overweighting of Fees Relative to Sample Returns

Moshe Levy

The Hebrew University

Richard Roll

California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 1430

July, 2017

Generalized Portfolio Performance Measures:

Optimal Overweighting of Fees Relative to Sample Returns

July 27, 2017

Moshe Levy* and Richard Roll**

Abstract

Performance measures such as alpha and the Sharpe ratio are typically based on sample returns net of fees. This implies the same weighting to sample returns and to fees. However, sample return parameters are noisy estimates of true parameters, while fees are known with certainty. Thus, intuition suggests that fees should be given more weight than sample returns. We formalize this intuition, and derive the optimal overweighting of fees. We show that the resulting generalized performance measures are better predictors of future net performance than the standard performance measures, and they better explain future fund flows.

Keywords: Mutual fund performance, alpha, Sharpe ratio, geometric mean, shrinkage, fees.

JEL Classification: G11.

*Jerusalem School of Business, The Hebrew University, Jerusalem 91905, Israel. 972-2-588-1329. mslm@huji.ac.il.

**California Institute of Technology, 1200 East California Blvd, Pasadena, CA 91125. 626-395-3890. rroll@caltech.edu.

1. Introduction

As of mid-2016, about 44% of U.S. households own mutual funds. The investment in mutual funds represents on average approximately one-fifth of households' financial assets.[1] There are thousands of mutual funds in the U.S. market alone, with assets totaling roughly $18.5 trillion. Thus, selecting a mutual fund is an important and complex decision faced by millions of investors.

Two key inputs for this decision are the fund's past performance and the fees it charges. Investors are, of course, interested in returns net of fees. Indeed, most reported performance measures are based on past net returns. This practice implicitly assigns the same weighting to sample gross returns and fees. For example, a fund with a gross sample alpha of 4% and fees of 3% has the same net alpha as a fund with a gross sample alpha of 2% and fees of 1%. However, there is a key difference between sample gross average returns (and other sample parameters) and fees: the sample returns are noisy estimates of the true parameters, while the fees are known with certainty.

Fees should be weighed more heavily than the sample returns. All else equal, one should prefer the fund with a sample alpha of 2% and fees of 1% over the fund with sample alpha of 4% and fees of 3%, because the 2% difference in fees is known with certainty, while the 2% difference in sample gross alphas could very well be due to sampling error. Another way to state this is that the sample alpha should be shrunk (in the sense of Bayes, or James-Stein) to the cross-sectional grand average, but fees, which are known, should not be. Thus, in contrast to the common practice of employing net returns in evaluating performance, we argue that one should treat gross sample returns and fees separately.

The degree to which one should overweight fees relative to sample returns depends on the informativeness of the sample returns. On the one extreme, if the sample return parameters perfectly reflect the true parameters, they should be given the same weight as the fees. On the other extreme, if the sample parameters contain no information at all about the true returns, they should be completely ignored, and the decision should

[1] See "Ownership of Mutual Funds, Shareholder Sentiment, and Use of the Internet, 2016", report by the Investment Company Institute, October 2016.

be based on the fees alone. In general, the larger the sampling error, the more one should shrink the sample parameters, implying that the more one should overweight fees relative to the sample returns. The main contribution of this study is the derivation of the optimal over-weighting of fees relative to the sample returns. This derivation yields simple generalized performance measures that are easy to implement.

When a fund constitutes the investor's entire risky investment, the relevant performance measure is the fund's Sharpe ratio. When an investor holds a fund as a part of a larger risky portfolio, the relevant performance measure is the fund's alpha with respect to the investor's entire portfolio.[2] Another well-known portfolio performance measure that has recently regained attention (Lo, Orr, and Zhang 2017, Levy 2017) is the geometric mean. The logic for overweighting fees applies to all three of these performance measures. We introduce the Generalized Sharpe ratio (GS), Generalized alpha (Galpha), and the Generalized Geometric Mean (GGM) as modifications of the standard performance measures that incorporate the optimal overweighting of fees.

The proposed generalized performance measures have bearing on one of the most central issues in the vast mutual fund literature: the persistence of performance. Most of the literature on performance persistence employs past performance and fund characteristics, fees typically among them, to predict future performance. Some of these studies find significant performance persistence (Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), Goetzmann and Ibbotson (1994), Brown and Goetzmann (1995), Wermers (1996), Elton, Gruber, Das and Blake (1996), Kosowski, Timmerman, Wermers, and White (2006), Barras, Scaillet, and Wermers (2010)). Others either don't find persistence (Sharpe (1966) and Jensen (1969)), or argue that after including the appropriate controls for momentum (Carhart (1997)), or the Fama-French (2015) profitability and investment factors, this persistence vanishes (Jordan and Riley 2015).

The present study contributes to this literature in two ways. First, while the above literature is empirically-driven, our analysis is theoretically-driven, and yields analytical

[2] Levy and Roll (2015) argue that even in this case, alpha is appropriate only if the investor considers infinitesimal adjustments to his portfolio. Otherwise, tilting the portfolio towards assets with positive alphas (and away from assets with negative alphas) is not very helpful, and one can do much better by constrained optimization on the assets' weights.

formulas for the generalized performance measures. Second, while studies that include fees as explanatory variables almost all find a negative coefficient for fees,[3] these cross-sectional regressions assign a single fee coefficient to all funds. In contrast, we argue that the optimal over-weighting of fees relative to the sample returns generally depends on fund-specifics, such as the number of return observations and the volatility of sample returns: the smaller the number of observations, and the higher the sample volatility, the more one should overweight fees.

2. Generalized Performance Measures

The basic idea underlying our generalized measures is that sample parameters should be shrunk towards their population means, but the fees are known, and should therefore be taken as is. This implies an under-weighting of the sample parameters relative to fees, or equivalently, an overweighting of the fees.

2.1 Generalized Alphas

Consider an investor who is interested in choosing the fund with maximal alpha net of fees. Alpha could be Jensen's (1969) original CAPM alpha, or the alpha with respect to any other factor model. For each fund *i*, the investor observes N_i independent sample observations, α_i^t, with a sample average $\hat{\alpha}_i$. These alphas are gross of fees. We assume that the α_i^t's of fund *i* are drawn from a normal distribution with a known standard deviation of σ_i, but an unknown mean, which is the fund's true population α_i (in practical and empirical applications σ_i will be estimated by its sample value $\hat{\sigma}_i$). The true α's are assumed to be distributed normally with mean μ_α and standard deviation σ_α. The fees of fund *i* are known and denoted by Fee_i, and they represent a fixed percentage of assets under management.

[3] See Sheng, Simutin, and Zhang (2017) for a recent exception.

The standard practice is to rank funds by their sample alphas net of fees, i.e. by:

$$\hat{\alpha}_i - Fee_i \,. \tag{1}$$

However, choosing a fund by this ranking does *not* yield the maximal expected ex-ante net alpha. Given a fund with gross sample alpha $\hat{\alpha}_i$, the expected ex-ante gross alpha is:

$$E(\alpha_i \mid \hat{\alpha}_i) = \frac{\dfrac{N_i}{\sigma_i^2}\hat{\alpha}_i + \dfrac{1}{\sigma_\alpha^2}\mu_\alpha}{\left(\dfrac{N_i}{\sigma_i^2} + \dfrac{1}{\sigma_\alpha^2}\right)}. \tag{2}$$

This is a straightforward application of Bayes Theorem (see, for example, eq.(24) in Murphy 2007). The expectation is "shrunk" towards the mean population alpha of μ_α.[4] Funds should be ranked by their expected net alphas, i.e. by the expected gross alpha given above, minus fees, which we define as the generalized alpha:

$$G_{alpha} \equiv \frac{\dfrac{N_i}{\sigma_i^2}\hat{\alpha}_i + \dfrac{1}{\sigma_\alpha^2}\mu_\alpha}{\left(\dfrac{N_i}{\sigma_i^2} + \dfrac{1}{\sigma_\alpha^2}\right)} - Fee_i \,. \tag{3}$$

Notice that this implies that the sample alpha is under-weighed relative to the fees (or conversely, the fees are overweighed relative to the sample alpha): the sample alpha is replaced by a weighted average of the sample alpha and the average alpha. In the extreme case where the estimation error, $\dfrac{\sigma_i^2}{N_i}$, is very large relative to the variation in the population alphas, σ_α^2, the sample alpha is completely ignored, and only the fees are taken into account. On the other extreme, if the estimation error is very small, the fraction in (3) converges to the sample alpha, and the standard net alpha (1)is obtained as a

[4] If the parameters $\mu_\alpha, \sigma_\alpha$, and σ_i are unknown, one can use the James-Stein (1956, 1961) shrinkage estimator, which converges to the Bayes estimator when the number of funds is large.

special case. In general, the larger the estimation error and the lower the variation in the population alphas, the more the sample $\hat{\alpha}_i$ should be under-weighed relative to the fees.

Empirically, we find that shrinking the alphas to zero or alternatively to the cross-sectional mean yield almost the same results.

2.2 Generalized Sharpe Ratio

Consider an investor who is interested in maximizing expected return for a given level of standard deviation, i.e. an investor who seeks to maximize the portfolio's Sharpe ratio. We make the following assumptions:

1. Investors can borrow and lend at the risk-free rate.
2. Investors know the true standard deviation σ_i for each fund, but not the mean return. (In practical and empirical applications σ_i will be estimated by its sample value $\hat{\sigma}_i$).
3. The true Sharpe ratios (gross of fees) are distributed normally across funds with mean μ_S and standard deviation σ_S.
4. Fund *i* has known and fixed fees Fee_i, which represents a fixed percentage of assets under management.

All returns are denoted in excess of the risk-free rate. The investor observes N independent gross return observations for each fund. Fund *i* has a sample mean return $\hat{\mu}_i$, The investor wishes to hold a portfolio with a standard deviation of σ_0. This can be accomplished by leveraging (or unleveraging) holdings in the selected fund to obtain a portfolio with the desired standard deviation σ_0. The investor's goal is to choose the fund that maximizes expected return, net of fees, given the standard deviation of σ_0. The value of σ_0 is arbitrary, as will become evident in what follows. If fund *i* is selected, the investor will invest a proportion $x = \frac{\sigma_0}{\sigma_i}$ of wealth in the fund (and a proportion $1-x$ in the risk-free asset), so that the portfolio will have the desired level of risk σ_0. This leveraged portfolio's sample gross mean is $\hat{\mu}_i^* = \frac{\sigma_0}{\sigma_i}\hat{\mu}_i$ (recall that returns are in excess of the risk-free rate), and its standard deviation is $\sigma_i^* = \sigma_0$, where the superscripts * indicate the parameters of the levered portfolio. The fees the investor will pay on this portfolio are $Fee_i^* = \frac{\sigma_0}{\sigma_i} Fee_i$ (for example, if the investor places \$50 in the risk-free asset and \$50 in a fund that charges fees of Fee_i =1% of assets under management, the fees of \$0.5, are 0.5% of the total portfolio, i.e. $Fee_i^* = 0.5\%$).

What is the expected net return for this portfolio? Let us first calculate the expected gross return, and then subtract the fees. Again, the expected ex-ante return should be shrunk toward the population average.[5] The standard error in estimating μ_i is $\frac{\sigma_i}{\sqrt{N}}$. For μ_i^* the standard error is therefore $x \cdot \frac{\sigma_i}{\sqrt{N}} = \frac{\sigma_0}{\sigma_i} \cdot \frac{\sigma_i}{\sqrt{N}} = \frac{\sigma_0}{\sqrt{N}}$. As the true Sharpe ratios are distributed normally across funds, with mean μ_S and standard deviation σ_S, for portfolios with standard deviation σ_0, the population means are distributed normally with mean $\sigma_0 \mu_S$ and standard deviation $\sigma_0 \sigma_S$. Thus, employing Bayes' rule, just as in the previous section, for a portfolio with sample gross mean $\hat{\mu}_i^*$ the expected gross ex-ante return is given by:

$$E\left(\mu_i^* \mid \hat{\mu}_i^*\right) = \left(\frac{\sigma_0^2 \sigma_S^2}{\sigma_0^2 \sigma_S^2 + \sigma_0^2 / N}\right)\hat{\mu}_i^* + \left(\frac{\sigma_0^2 / N}{\sigma_0^2 \sigma_S^2 + \sigma_0^2 / N}\right)\sigma_0 \mu_S = \left(\frac{1}{1 + 1/N\sigma_S^2}\right)\hat{\mu}_i^* + \left(\frac{1}{1 + N\sigma_S^2}\right)\sigma_0 \mu_S \tag{4}$$

[5] Technically, the shrinkage should be done only when comparing three or more funds (in conformity to the James-Stein method.)

In this case the shrinkage of $\hat{\mu}_i^*$ is towards the mean expected return across all portfolios with standard deviation σ_0, i.e. towards $\sigma_0 \mu_S$. The expected net return is:

$$E\left(\mu_i^* \mid \hat{\mu}_i^*\right) - Fee_i^* = \left(\frac{1}{1+1/N\sigma_S^2}\right)\hat{\mu}_i^* + \left(\frac{1}{1+N\sigma_S^2}\right)\sigma_0 \mu_S - Fee_i^* . \qquad (5)$$

In terms of the parameters of the original unlevered fund, the expected net return is:

$$E\left(\mu_i^* \mid \hat{\mu}_i^*\right) - Fee_i^* = \left(\frac{1}{1+1/N\sigma_S^2}\right)\frac{\sigma_0}{\sigma_i}\hat{\mu}_i + \left(\frac{1}{1+N\sigma_S^2}\right)\sigma_0 \mu_S - \frac{\sigma_0}{\sigma_i} Fee_i . \qquad (6)$$

Note that the second term is just a constant that is the same for all funds, and it is therefore irrelevant for the ranking of funds, and can thus be ignored.[6] Similarly, the term σ_0 that multiplies all terms is irrelevant for the ranking. Thus, funds should be ranked by the following Generalized Sharpe ratio:

$$GS_i = \frac{\left(\frac{1}{1+1/N\sigma_S^2}\right)\hat{\mu}_i - Fee_i}{\sigma_i} . \qquad (7)$$

This is very similar to the standard Sharpe ratio, except for the term $\left(\frac{1}{1+1/N\sigma_S^2}\right)$ multiplying the sample mean $\hat{\mu}_i$ (recall that returns are in excess of the risk free rate, i.e. the standard Sharpe ratio is: $S_i = \frac{\hat{\mu}_i - Fee_i}{\sigma_i}$). This term is smaller than 1, implying the under-weighting of the sample mean relative to the known fees. If the number of observations, *N*, or the variation in the true Sharpe ratios across funds, σ_S, become very large, then the Generalized Sharpe ratio converges to the standard Sharpe ratio.

[6] In the case that the number of return observations is different across funds, this term is no longer constant, and the generalized Sharpe ratio becomes: $GS_i = \frac{\left(\frac{1}{1+1/N_i\sigma_S^2}\right)\hat{\mu}_i - Fee_i}{\sigma_i} + \left(\frac{1}{1+N_i\sigma_S^2}\right)\mu_S$.

2.3 Generalized Geometric Mean

Choosing the fund with the highest geometric mean (GM) has been suggested decades ago as the "growth optimum" strategy that over the long run almost surely yields a higher terminal wealth than investment in any other fund (Kelly 1956, Latane 1959, Markowitz 1976). Recently, Levy (2017) suggests that the GM provides a very good measure of portfolio performance, even if the horizon is not long. He shows that in a realistic setting with limited borrowing the GM yields a ranking of funds that is more aligned with investors' expected utilities than the alignment obtained when ranking funds by the Sharpe ratio or by alpha. This is true for a wide range of preferences, and for horizon as short as one month. Another advantage of the GM as a performance measure is that while the ranking of funds by alphas or by Sharpe ratios may change with the investment horizon, even if returns are i.i.d, (Levy 1972, Levhari and Levy 1977), the ranking by the GM is invariant to the investment horizon. Thus, it is of interest to develop the generalized GM performance measure.

Consider an investor who is focused on maximizing her portfolio's geometric mean, net of fees. Assume that the *total* returns of fund *i*, gross of fees, R_i^{gross}, are distributed log-normally, i.e. $\log\left(\tilde{R}_i^{gross}\right) \sim N\left(g_i^{gross}, \sigma_i\right)$, and R=1+gross rate of return. Note that the log of the gross geometric mean is $g_i^{gross} = E\left(\log\left(\tilde{R}_i^{gross}\right)\right) = \log\left(GM_i^{gross}\right)$. We assume that the true log-gross-geometric-mean, g^{gross}, is distributed normally across funds, with mean μ_g and standard deviation σ_g. The investor does not know the true parameter g_i^{gross}, but observes N_i gross return observations for fund *i*, with a sample gross log-geometric-mean $\hat{g}_i^{gross} = \frac{1}{N_i}\sum_{j=1}^{N_i}\log\left(R_i^{gross,\,t-j}\right)$, where *t* is the current time, and $R_i^{gross,t-j}$ indicates the gross total return of fund *i* at time *t-j*. As before, it is assumed that σ_i is known (and estimated by $\hat{\sigma}_i$ in practical applications).

The fees are known. In the present context it is convenient to denote the fees in terms of the assets under management at the *end* of the period, such that: $R_i^{net} = R_i^{gross} \cdot (1 - Fee_i)$.[7] This implies that:

$$\log\left(GM_i^{net}\right) = \log\left(GM_i^{gross}\right) - \log(1 - Fee_i), \text{ or:}$$

$$g_i^{net} = g_i^{gross} - \log(1 - Fee_i). \tag{8}$$

We assume that the investor is interested in selecting the fund that has the maximal expected value of g^{net}, i.e. the fund that has the highest expected log(*GM*), net of fees.

The standard error in the estimation of $\hat{g}_i^{gross}$ is $\frac{\sigma_i}{\sqrt{N_i}}$. The standard deviation of the true values of g^{gross} across the population of funds is distributed normally with mean μ_g and standard deviation σ_g. Thus, given the observation $\hat{g}_i^{gross}$, by Bayes' rule the expected ex-ante g_i^{gross} is given by:

$$E\left(g_i^{gross} \mid \hat{g}_i^{gross}\right) = \left(\frac{\sigma_g^2}{\sigma_g^2 + \sigma_i^2 / N}\right)\hat{g}_i^{gross} + \left(\frac{\sigma_i^2 / N}{\sigma_g^2 + \sigma_i^2 / N}\right)\mu_g. \tag{9}$$

The Generalized Geometric Mean (GGM) is given by the expected net g, i.e. by:

[7] This definition is slightly different than the one used in the previous sections, where it is assumed that the fees are in terms of the assets under management at the *beginning* of the period, i.e. that $R_i^{net} = R_i^{gross} - Fee_i$. However, for practical purposes both definitions are very similar. For example, if one invests \$100, the gross return is 5%, and the fees are 2%, if these fees are defined in terms of the assets in the beginning of the period, the net end-of-period value is \$100(1.05-0.02)=\$103. If, alternatively, the fees are defined in terms of the assets of the end of the period, the net end-of-period value is \$100(1.05(1-0.02))=\$102.9.

$$GGM_i = E\left(g_i^{net} \mid \hat{g}_i^{gross}\right) = \left(\frac{\sigma_g^2}{\sigma_g^2 + \sigma_i^2 / N}\right)\hat{g}_i^{gross} + \left(\frac{\sigma_i^2 / N}{\sigma_g^2 + \sigma_i^2 / N}\right)\mu_g - \log\left(1 - Fee_i\right) \tag{10}$$

, or:

$$GGM_i = \left(\frac{1}{1 + \sigma_i^2 / N\sigma_g^2}\right)\hat{g}_i^{gross} + \left(\frac{1}{1 + N\sigma_g^2 / \sigma_i^2}\right)\mu_g - \log\left(1 - Fee_i\right). \tag{11}$$

As in the case of the generalized alpha and the generalized Sharpe ratio, here too, the observation of $\hat{g}_i^{gross}$ is underweighted relative to the fees: $\hat{g}_i^{gross}$ is shrunk towards the population mean μ_g, while the known fees are taken as is.

3. Empirical Analysis

3.1 Data

Our data are from the CRSP survivorship-bias-free mutual fund data set. We employ domestic U.S. equity funds. We include only funds with total net asset value exceeding \$20 million, and investment of between 70% to 130% of the fund's assets in common equity. As total net assets (needed for calculating fund flows) are reported monthly on a regular basis since 1992, we employ the January 1992 – December 2016 sample period.

3.2 Prediction of Future Performance

The preceding theoretical analysis suggests that the generalized performance measures are better estimates of the funds' true performance than the standard measures. In order to examine this empirically, we analyze the correlation between in-sample performance and out-of-sample performance as a function of the under-weighting. Note that the under-weighting of the sample parameters depends on σ_α, σ_S and σ_g, which are generally unknown. Our analysis allows us to empirically find the under-weighting that works best. We take an in-sample period of 36 months, and an out-of-sample period of the next 36 months. The out-of-sample performance is calculated from the out-of-sample net returns. The in-sample performance is calculated by the generalized performance measures (eq. 3, with mean alpha assumed to be zero, 7 and 11), for different values of the under-weighting of the sample parameters.

Figure 1 shows this relation for alphas. σ_α determines the under-weighting of the sample alpha. $1/\sigma_\alpha = 0$ ($\sigma_\alpha = \infty$) implies no shrinking of the sample alpha at all – in this case the generalized alpha reduces to the standard alpha calculated with net returns. The other extreme of $1/\sigma_\alpha = \infty$ ($\sigma_\alpha = 0$) implies shrinking all sample alphas to zero, i.e. ranking funds only according to their fees. Figure 1 shows the correlation between the in-sample generalized alpha and the out-of-sample alpha as a function of $1/\sigma_\alpha$. The three panels correspond to three different versions of alpha: A: CAPM alpha, B: Four-Factor alpha (the 3 Fama-French factors + momentum), and C: alphas of the Fama-French (2015) Five-Factor model.

In all three cases the generalized alphas offer an improvement over their standard counterparts. For the CAPM alpha, the highest correlation is obtained for a value of approximately $1/\sigma_\alpha^{CAPM} = 100$, i.e. $\sigma_\alpha^{CAPM} = 0.01$. This value implies that for a typical fund with a standard deviation of monthly returns, σ_i of, say, 5%, the sample alpha is under-weighed by a factor of 0.59, because eq.(3) becomes:

$$G_{alpha} \equiv \left(\frac{1}{1+\frac{1}{N_i}\frac{\sigma_i^2}{\sigma_\alpha^2}} \right) \hat{\alpha}_i - Fee_i = \left(\frac{1}{1+\frac{1}{36}\frac{0.05^2}{0.01^2}} \right) \hat{\alpha}_i - Fee_i = 0.59 \cdot \hat{\alpha}_i - Fee_i .$$

For a less volatile fund, with monthly σ_i of 3%, the under-weighting is by a factor of 0.8:

$$G_{alpha} \equiv \left(\frac{1}{1+\frac{1}{N_i}\frac{\sigma_i^2}{\sigma_\alpha^2}} \right) \hat{\alpha}_i - Fee_i = \left(\frac{1}{1+\frac{1}{36}\frac{0.03^2}{0.01^2}} \right) \hat{\alpha}_i - Fee_i = 0.8 \cdot \hat{\alpha}_i - Fee_i .$$

For the Five-Factor alpha the maximal correlation is obtained for $1/\sigma_\alpha^{5F} = 125$, implying more under-weighting. For instance, for a fund with $\sigma_i = 5\%$ the sample alpha is under-weighed by a factor of 0.48. For the Four-Factor alpha one should apply even more under-weighting. In this case the maximal correlation is obtained for a value of approximately $1/\sigma_\alpha^{4F} = 180$, implying that for a fund with $\sigma_i = 5\%$ the sample alpha is under-weighed by a factor of 0.31.

When considering the Sharpe ratio and the GM, the correlation between the in-sample standard measure (with no under-weighting) and the out-of-sample measure in our sample turns out to be negative (-0.37 in the case of the Sharpe ratio, and -0.45 in the case of the GM). These negative correlations are likely due to the negative correlation of the average market returns over the same period – in our sample the correlation between the average monthly return on the market in a 36-month period and the average return in the next 36-month period is -0.36). This problem does not affect the alphas, as they control for the market factor. The negative correlations imply a corner solution – ignore the sample measures altogether and rank funds only according to their fees. However, we believe that this situation may be rather specific to the sample period employed. Unfortunately, the sample period cannot be changed much due to the availability fee data, and negative correlations also appear when other reasonable horizons are employed instead of 36 months. Thus, to estimate the optimal under-weighting for the Sharpe ratio and the GM we take a modified approach.

Every month, we sort funds into 100 groups, sorted by their in-sample generalized performance. Thus, the composition of these 100 groups generally changes from month to month. For each of the groups, we calculate the average in-sample generalized performance over the preceding 36 months, and the average out-of-sample net performance over the next 36 months. The averaging is across all funds in the group at a given month, and across all months in the sample, i.e. for each one of the 100 groups we have one in-sample average performance, and one out-of-sample average performance. The averaging across time neutralizes the negative correlation problem, yet still allows us to maintain the relation between in-sample and out-of-sample performance. This analysis is conducted for different levels of under-weighting of the sample parameters, and we examine the correlation as a function of the under-weighting. Figure 2 shows the results: Panel A corresponds to the Sharpe ratio, and Panel B to the geometric mean.

For the Sharpe ratio, the correlation is maximized at the value of $1/\sigma_S = 16$, or $\sigma_S = 0.0625$. This value of σ_S is much larger than the typical value of σ_α (for the

CAPM alpha, the maximal correlation is obtained for $\sigma_{\alpha}^{CAPM} = 0.01$). This is reasonable, as the monthly alphas are in the order of a few percent, while the monthly Sharpe ratio is in the order of about 0.2. For N=36 months, this value implies an under-weighting of the sample mean by a factor of about 0.12 (see eq.7). This rather dramatic under-weighting conforms with the notorious difficulty of estimating mean returns.

The Generalized Geometric Mean (GGM) requires two global parameters: the mean and standard deviation of the log(GM), μ_g and σ_g, respectively (see eq. 11). For μ_g we take the average in our sample, which is 0.0064. We then examine the correlation between the in-sample GGM and the out-of-sample GM as a function of σ_g. The maximal correlation is obtained for a value of about $1/\sigma_g = 120$, or $\sigma_g = 0.0083$. This value is similar to those obtained for alphas. For a typical fund with $\sigma_i = 5\%$, it implies that the gross sample $\hat{g}_i^{gross}$ is shrunk toward the average μ_g as:

$$\left(\frac{1}{1+\left(0.05^2/36\cdot 0.0083^2\right)}\right)\hat{g}_i^{gross} + \left(\frac{1}{1+\left(36\cdot 0.0083^2/0.05^2\right)}\right)\mu_g = 0.5\cdot\hat{g}_i^{gross} + 0.5\cdot\mu_g .$$

3.3 Fund Flows

It has been well-documented that fund flows react to past performance (Chevalier and Ellison 1997, Elton, Gruber, and Busse 2004, Barber, Odean, and Zheng 2005, Spiegel and Zhang 2013, and Fulkerson and Riley 2016). If investors intuitively understand that fees should be weighed more heavily than sample parameters, we would expect flows to be more closely related to the generalized performance measures than to their standard counterparts. The purpose of this section is to examine this question empirically.

Fund flows are calculated as usual:

$$Flow_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1}\left(1 + r_{i,t-1}\right)}{TNA_{i,t-1}}, \tag{12}$$

where $TNA_{i,t}$ denotes the total net assets of fund i at time t, and $r_{i,t}$ is the fund's rate of return between time *t-1* and *t*.

We regress the fund flow at time *t* on the fund's past performance measure (*PPM*) during the period (*t*-*T*, *t*-1). Fund performance is taken either the standard measure (alpha, Sharpe, GM), net of fees, or the corresponding generalized measure (G_{alpha}, GS, GGM). For alphas, we examine all three versions: CAPM alpha, Four-factor alpha, and Five-Factor alpha. Specifically, we run the regression:

$$Flow_{i,t} = a + b \cdot PPM_{i,t-T:t-1} + \sum_{k=1}^{K} c_k \cdot control_{i,k} + \varepsilon_{i,t} \tag{13}$$

We use the following controls: the standard deviation of monthly returns over (*t*-*T*, *t*-1), log(fund size), fund age (in years), manager tenure (in years), and turnover. The in-sample period is taken, as before, as 36 months (T=36), and the flow is measured over the next month. For the generalized performance measures, we take the optimal under-weighting as found in the preceding section.

The results are given in Table 1. Note that the generalized performance measures are typically smaller than their standard counterparts, because of the shrinkage, and therefore, their coefficients (and standard errors) are larger. The more relevant results are their significance (t-values, given in parentheses). In all cases, the generalized performance measure is more significant than its standard counterpart. This suggests that investors do indeed realize that fees should be overweighed relative to the sample parameters, and are more sensitive to the generalized performance measures than to the standard measures.

4. Conclusions

An effective tool for fund selection is of great importance to investors. In practice, virtually all performance measures are based on past returns. As investors are interested in returns net of fees, the standard practice is to calculate performance measures based on past net returns. This implicitly assigns the same weight to past returns and to fees: a fund with a gross sample average return of 4% and fees of 3% has the same net average return as a fund with a gross sample average return of 2% and fees of 1%.

The point of this paper is very simple: past returns and fees should be treated differently: past returns are noisy estimates and should be "shrunk" to the cross-sectional average, while fees are known with certainty, and should be taken as is. This implies that fees should be weighed more heavily than past returns. This study formalizes this idea, and captures it by suggesting generalizations of the three main portfolio performance measures: alpha, the Sharpe ratio, and the geometric mean.

These generalized performance measures better predict future performance, and thus constitute an improvement over the standard measures. Furthermore, future fund flows are better explained by the generalized performance measures than by their standard counterparts, suggesting that investors intuitively understand that fees should be overweighed relative to sample returns. It is our hope that the simple generalized performance measures suggested here will be adopted by investors and by fund rating agencies, as superior alternatives to the standard measures currently employed.

References

Barber, Brad M., Terrance Odean, and Lu Zheng. "Out of sight, out of mind: The effects of expenses on mutual fund flows." *The Journal of Business* 78, (2005): 2095-2120.

Barras, Laurent, Olivier Scaillet, and Russ Wermers. "False discoveries in mutual fund performance: Measuring luck in estimated alphas." *The Journal of Finance* 65, no. 1 (2010): 179-216.

Bayes, Thomas and Richard Price, "An essay towards solving a problem in the doctrine of chance" *Philosophical Transactions of the Royal Society of London*. 53 (1763): 370–418.

Brown, Stephen J., and William N. Goetzmann. "Performance persistence." *The Journal of Finance* 50, no. 2 (1995): 679-698.

Carhart, Mark M. "On persistence in mutual fund performance." *The Journal of Finance* 52, no. 1 (1997): 57-82.

Chevalier, Judith, and Glenn Ellison. "Risk taking by mutual funds as a response to incentives." *Journal of Political Economy* 105, (1997): 1167-1200.

Elton, Edwin J., Martin J. Gruber, and Jeffrey A. Busse. "Are investors rational? Choices among index funds." *The Journal of Finance* 59, (2004): 261-288.

Elton, Edwin J., Martin J. Gruber, Sanjiv Das, and Christopher R. Blake. "The persistence of risk-adjusted mutual fund performance." *Journal of Business* 69,(1996): 133-157.

Fama, Eugene F., and Kenneth R. French. "Luck versus skill in the cross-section of mutual fund returns." *The Journal of Finance* 65, no. 5 (2010): 1915-1947.

Fama, Eugene F., and Kenneth R. French. " A five-factor asset pricing model". *Journal of Financial Economics*, 116 (2015), pp. 1–22.

Fulkerson, Jon A., and Timothy B. Riley. "Do Investors Chase Performance or Skill? Evidence from Mutual Fund Flows." SSRN: https://ssrn.com/abstract=2910705 (2016).

Goetzmann, William N., and Roger G. Ibbotson. "Do winners repeat?." *The Journal of Portfolio Management* 20, no. 2 (1994): 9-18.

Grinblatt, Mark, and Sheridan Titman. "The persistence of mutual fund performance." *The Journal of Finance* 47, no. 5 (1992): 1977-1984.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser. "Hot hands in mutual funds: Short-run persistence of relative performance, 1974–1988." *The Journal of Finance* 48, no. 1 (1993): 93-130.

James, William, and Charles Stein. "Estimation with quadratic loss." In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, (1961): 361-379.

Jensen, Michael C. "Risk, the pricing of capital assets, and the evaluation of investment portfolios." *The Journal of Business* 42, no. 2 (1969): 167-247.

Jordan, Bradford D., and Timothy B. Riley. "Volatility and mutual fund manager skill." *Journal of Financial Economics* 118, no. 2 (2015): 289-298.

Kosowski, Robert, Allan Timmermann, Russ Wermers, and Hal White. "Can mutual fund “stars” really pick stocks? New evidence from a bootstrap analysis." *The Journal of Finance* 61, no. 6 (2006): 2551-2595.

Latane, H. A. “Criteria for choice among risky ventures”. *Journal of Political Economy*, no.67, (1959): 144-155.

Levhari, D., & Levy, H. “The capital asset pricing model and the investment horizon”. *The Review of Economics and Statistics*, (1977): 92-104.

Levy, H. “Portfolio performance and the investment horizon”. *Management Science*, no. 18, (1972): B-645.

Levy, Moshe. "Measuring Portfolio Performance: Sharpe, Alpha, or the Geometric Mean?." *Journal of Investment Management*, forthcoming (2017).

Levy, Moshe, and Richard Roll. "Seeking alpha? It’s a bad guideline for portfolio optimization." *The Journal of Portfolio Management* 42, (2015): 107-112.

Lo, Andrew W. and Orr, Allen and Zhang, Ruixun, “The growth of relative wealth and the Kelly criterion” (2017). Working paper available at SSRN: https://ssrn.com/abstract=2900509.

Murphy, Kevin P. "Conjugate Bayesian analysis of the Gaussian distribution." *def* 1, no. 2σ2 (2007): 16.

Sharpe, William F. "Mutual fund performance." *The Journal of Business* 39, no. 1 (1966): 119-138.

Sheng, Jinfei, Mikhail Simutin, and Terry Zhang. "Cheaper is not better: on the superior performance of high-fee mutual funds.” SSRN: 2912511 (2017)..

Spiegel, Matthew, and Hong Zhang. "Mutual fund risk and market share-adjusted fund flows." *Journal of Financial Economics* 108, no. 2 (2013): 506-528.

Stein, Charles. "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution." In *Proceedings of the Third Berkeley symposium on mathematical statistics and probability*, (1956): 197-206.

Table 1

Fund flow at time t is regressed on the fund past performance during the period (t-T, t-1). As fund past performance measure (PPM) we take either the standard measures (alpha, Sharpe, geometric mean), net of fees, or the corresponding generalized measures, G_{alpha}, GS, GGM, as given by equations 3, 7, and 11. We regress:

$$Flow_{i,t} = a + bPPM_{i,t-T:t-1} + \sum_{k=1}^{K} c_k \cdot control_{i,k} + \varepsilon_{i,t}$$

The following controls are employed: the standard deviation of monthly returns over (t-T, t-1), log(fund size), fund age (in years), manager tenure (in years), and turnover. The in-sample period is taken as 36 months (T=36), and the flow is measured over the next month. For the generalized performance measures, we take the optimal under-weighting as reported is Section 3.2. Numbers in parentheses are the t-values. ** indicates significance at the 5% level, *** indicates significance at the 1% level.

Performance measure:	**Sharpe Ratio**		**CAPM α**		**Four-Factor α**		**Five Factor α**		**Geometric Mean**	
	Standard SR	Generalized SR	Standard α	Generalized α	Standard α	Generalized α	Standard α	Generalized α	Standard GM	Generalized GM
constant	1.17 (3.08***)	1.54 (4.40***)	1.41 (4.05***)	1.44 (4.13***)	1.44 (4.15***)	1.50 (4.29***)	1.43 (4.30***)	1.57 (4.46***)	1.25 (3.46***)	1.12 (2.93***)
Past performance measure	0.63 (1.72)	7.87 (3.11***)	22.83 (1.62)	74.41 (2.15**)	16.05 (0.97)	68.79 (1.70)	30.76 (2.18**)	92.48 (2.54**)	20.48 (1.93)	33.29 (2.00**)
σ	14.19 (1.92)	10.30 (1.46)	12.82 (1.78)	14.18 (1.95)	11.14 (1.57)	11.90 (1.67)	10.58 (1.57)	11.66 (1.64)	11.37 (1.61)	13.02 (1.81)
Log(TNA)	-0.29 (-7.69***)	-0.32 (-8.12***)	-0.30 (-7.70***)	-0.30 (-7.83***)	-0.29 (-7.60***)	-0.30 (-7.73***)	-0.29 (-7.79***)	-0.31 (-7.91***)	-0.30 (-7.73***)	-0.30 (-7.76***)
Fund age	0.003 (0.62)	0.004 (0.67)	0.004 (0.69)	0.004 (0.73)	0.004 (0.65)	0.004 (0.69)	0.004 (0.71)	0.004 (0.71)	0.004 (0.66)	0.004 (0.66)
Manager Tenure	0.001 (0.13)	0.002 (0.18)	0.002 (0.21)	0.003 (0.25)	0.002 (0.18)	0.002 (0.21)	0.002 (0.23)	0.002 (0.21)	0.002 (0.20)	0.002 (0.21)
Turnover	-0.24 (-2.35**)	-0.21 (-2.01**)	-0.26 (-2.46**)	-0.25 (-2.40**)	-0.26 (-2.47**)	-0.25 (-2.42**)	-0.25 (-2.54**)	-0.26 (-2.49**)	-0.25 (-2.45**)	-0.25 (-2.42**)

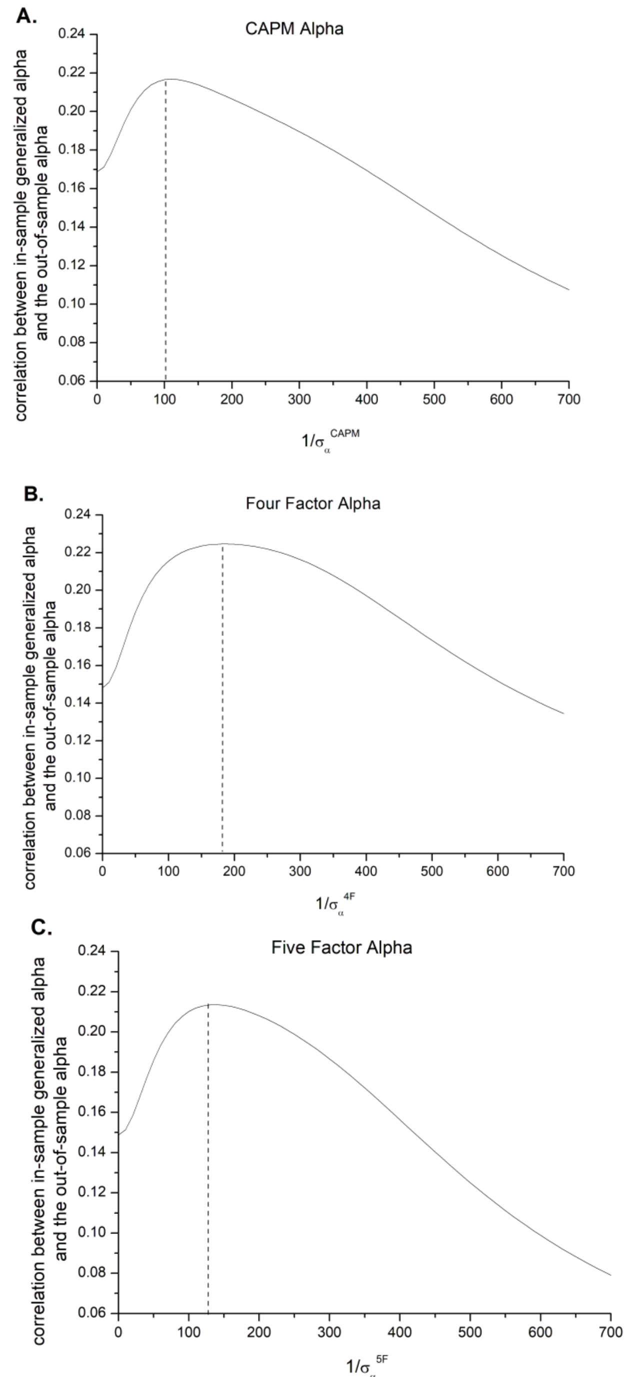

Figure 1: The correlation between in-sample generalized alpha and the out-of-sample net alpha as a function of the amount of shrinkage, as measured by $1/\sigma_\alpha$. We examine this relation for three versions of alpha: CAPM alpha (Panel A), Four-Factor alpha (Panel B), and the recently suggested Five-Factor alpha (Panel C). $1/\sigma_\alpha$ =0 implies no shrinkage at all, i.e. the generalized alphas reduce to the standard alphas.

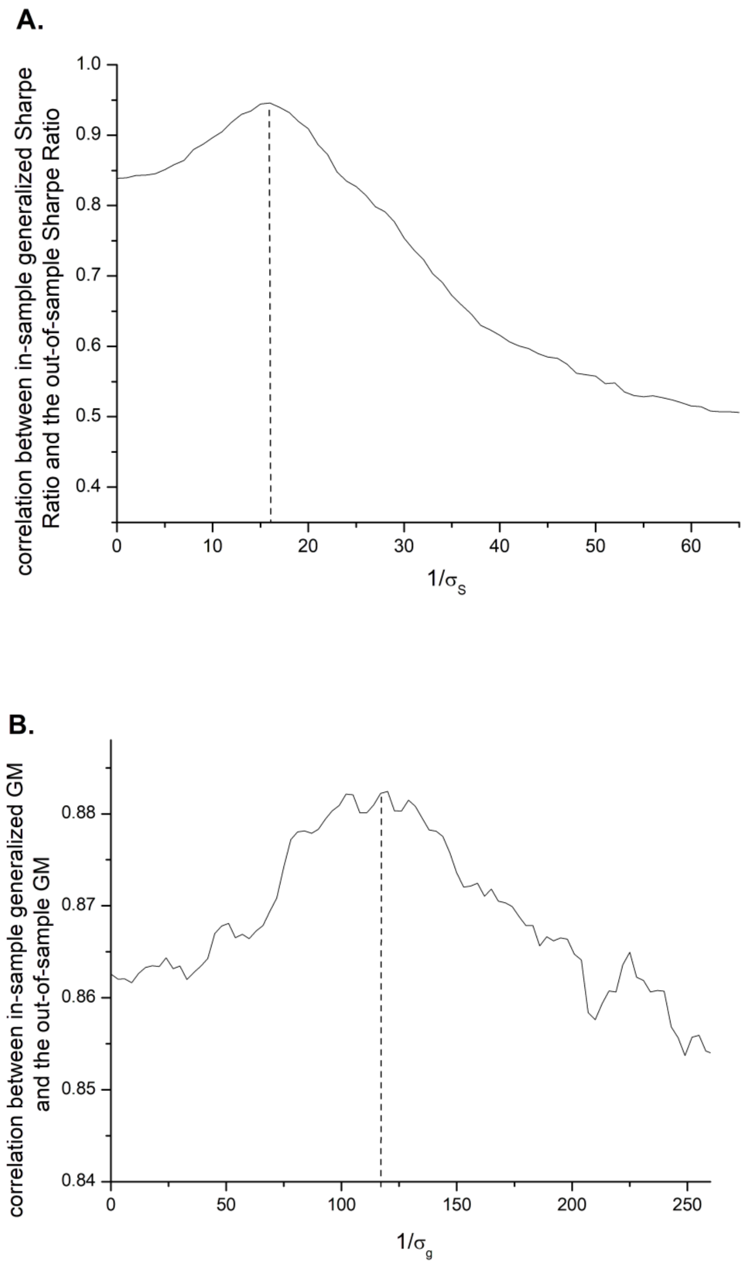

Figure 2: The correlation between in-sample generalized performance measures and the out-of-sample performance as a function of the amount of shrinkage, as measured by $1/\sigma_S$ in the case of the generalized Sharpe ratio (eq. 7), shown in Panel A, and by $1/\sigma_g$ in the case of the generalized geometric mean (eq. 11), shown in Panel B. $1/\sigma=0$ implies no shrinkage at all, i.e. the generalized measure reduces to the standard performance measure.